AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. (Currently Amended) A method of performing speaker verification to determine whether a speaker is a registered speaker, the method comprising:

a) obtaining an array ~~a plurality~~ of frames of compressed audio formants representing the speaker uttering a predetermined pass phrase, each frame within the array including:

i) energy data and pitch data characterizing the residue of the speaker uttering the predetermined pass phrase; and

ii) a plurality of formant coefficients characterizing the resonance of the speaker uttering the predetermined pass phrase; and

b) performing a time domain normalization of the array of frames of compressed audio formants to a sample array of frames of compressed audio formants such that such that the two arrays are of an equal quantity of frames;

c) determining whether the speaker is the registered speaker ~~verifying the identity of the speaker~~ by:

generating an array of discrepancy values, each discrepancy value representing the difference between ~~matching at least~~ one of: i) an energy ~~data~~ value; ii) a pitch ~~data~~ value; and iii) a formant ~~coefficients~~ coefficient value of a frame of the array and a corresponding energy value; ii) pitch value; and iii) formant coefficient value of a corresponding frame in the sample array; and

determining whether the array of discrepancy values is within a predetermined threshold. ~~in the frames to at least one of energy, pitch, and formant coefficients of a plurality of sample frames stored in memory.~~

2. (Currently Amended) The method of performing speaker verification of claim 22, ~~1,~~ wherein the step of obtaining an array of frames of compressed audio

formants includes receiving the frames of compressed audio formants from a remote Internet telephony device.

3. (Currently Amended) The method of performing speaker verification of claim 2, wherein the step of obtaining an array of frames of compressed audio formants from the remote Internet telephony device comprises receiving audio input of the speaker uttering the pass phrase from a microphone, and digitizing the audio input, converting the digitized audio input to a sequence of frames of compressed audio formants. ~~, further compressing the sequence of frames of compressed audio formants to generate compressed audio data packets, and sending the compressed audio data packets from the remote Internet telephony device.~~

4. (Canceled)

5. (Currently Amended) A method of determining whether a speaker is a registered speaker, the method comprising:

a) obtaining compressed audio formants for each frame of an array of frames representing the speaker uttering a predetermined pass phrase: ~~, the compressed audio formants including:~~

~~i) energy data and pitch data characterizing the residue of the speaker uttering the predetermined pass phrase;~~

~~ii) formant coefficients characterizing the resonance of the speaker uttering the predetermined pass phrase;~~

b) performing a time domain normalization of the array to a sample array of frames stored in a memory and representing the registered speaker uttering the predetermined pass phrase to decimate a portion of the frames of the larger of the two arrays such that the two arrays, after decimation, are of an equal quantity of frames, the portion of the frames to be decimated being selected by:

selecting a plurality of audio formant decimation groups, each audio formant decimation group being a selection of frames from the larger of the two arrays which, if decimated, yields the best alignment between a formant coefficient value of each frame of each the array and the corresponding formant coefficient value of each frame of the sample array; and

determining a decimation group of frames from the larger of the two arrays, the decimation group being a quantity of frames equal to the quantity of frames to be decimated and being the frames which are selected by weighted average from each of the audio formant decimation groups;

c) generating an array of discrepancy values, each discrepancy value representing the difference between one of an audio formant value of a frame of the array and a corresponding audio formant value of a corresponding frame of the sample array; and

d) determining that the remote speaker is the registered speaker if the array of discrepancy values is within a predetermined threshold. ~~determining whether the speaker is the registered speaker by matching at least one of energy, pitch, and formant coefficients from the compressed audio formants to predetermined combinations of at least one of energy, pitch, and formant coefficients of sample compressed audio formants known to represent the registered speaker.~~

6. (Currently Amended) The method of determining whether a speaker is a registered speaker of claim 23, ~~5,~~ wherein the step of obtaining compressed audio formants includes obtaining the compressed audio formants from a remote location and sending the compressed audio formants from the remote location.

7. (Original) The method of determining whether a speaker is a registered speaker of claim 6, wherein the step of obtaining compressed audio formants at a remote location includes receiving audio input of the speaker uttering the pass phrase from a microphone, digitizing the audio input, and compressing the digitized audio input to generate compressed audio formants.

Claims 8 – 9 (Cancelled)

10. (Currently Amended) A speaker verification server for determining whether ~~verifying the identity of~~ a remote speaker is a registered speaker, the server comprising:

a) a network interface for receiving, via a packet switched network, compressed audio formants for each frame of an array of frames ~~via a packet switched network~~ representing the ~~a~~ remote speaker uttering a predetermined pass phrase as audio input to a remote telephony client;

b) a database storing compressed audio formants for each frame of a sample array of frames ~~a plurality of compressed audio formant samples, each~~ representing the ~~a~~ registered speaker uttering ~~a registered~~ the predetermined pass phrase as audio input; and

c) a verification application operatively coupled to each of the network interface and the database for comparing the compressed audio formants of the array of frames to the compressed audio formants of the sample array of frames ~~representing the remote speaker to a compressed audio formant sample~~ to determine whether the remote speaker is the registered speaker by:

performing a time domain normalization of the array to the sample array such that such that the two arrays are of an equal quantity of frames;

generating an array of discrepancy values, each discrepancy value representing the difference between one of an audio formant value of a frame of the array and a corresponding audio formant value of a corresponding frame of the sample array; and

determining that the remote speaker is the registered speaker if the array of discrepancy values is within a predetermined threshold.

11. (Currently Amended) The speaker verification server of claim 24, ~~10,~~ wherein the compressed audio formants include energy data and pitch data

characterizing the residue of the speaker uttering the predetermined pass phrase and formant coefficients characterizing the resonance of the speaker uttering the predetermined pass phrase; and each frame ~~compressed audio formant sample~~ includes an energy value and a pitch value ~~data~~ characterizing the residue of the registered speaker uttering the registered pass phrase and formant coefficient values ~~coefficients~~ characterizing the resonance of the registered speaker uttering the registered pass phrase.

12. (Currently Amended) The speaker verification server of claim 11, wherein the verification application determines the decimation group of frames by:

selecting a pitch decimation group of frames from the larger of the two arrays, the pitch decimation group being the selection of frames which, if decimated, yields the best alignment between the pitch values of the two arrays after decimation;

selecting an energy decimation group of frames from the larger of the two arrays, the energy decimation group being the selection of frames which, if decimated, yields the best alignment between the energy values of the two arrays after decimation;

selecting a plurality of formant coefficient decimation groups, each formant coefficient decimation group being a selection of frames from the larger of the two arrays which, if decimated, yields the best alignment between the formant coefficient values of the two after decimation; and

selecting frames from the larger of the two arrays for the decimation group by weighted average from the pitch decimation group, the energy decimation group, and each formant coefficient decimation group. ~~determines whether the at least one of energy, pitch, and formant coefficients from the compressed audio formants is similar to the at least one of the energy, pitch, and formant coefficients of the sample.~~

Claims 13 - 21 (Canceled)

22. (New Claim) The method of performing speaker verification of claim 1, wherein performing a time domain normalization comprises:

comparing the quantity of frames in the array with the quantity of frames in the sample array to determine the quantity of frames to be decimated from the larger of the two arrays such that the two arrays are of an equal quantity of frames;

selecting a pitch decimation group of frames from the larger of the two arrays, the pitch decimation group being the selection of frames which, if decimated, yields the best alignment between the pitch values of the two arrays after decimation;

selecting an energy decimation group of frames from the larger of the two arrays, the energy decimation group being the selection of frames which, if decimated, yields the best alignment between the energy values of the two arrays after decimation;

selecting a plurality of formant coefficient decimation groups, each formant coefficient decimation group being a selection of frames from the larger of the two arrays which, if decimated, yields the best alignment between the formant coefficient values of the two arrays after decimation; and

determining a decimation group of frames from the larger of the two arrays, the decimation group being a quantity of frames equal to the quantity of frames to be decimated and being the frames which are selected by weighted average from the pitch decimation group, the energy decimation group, and each formant coefficient decimation group; and

decimating the decimation group of frames from the larger of the two arrays.

23. (New Claim) The method of determining whether a speaker is a registered speaker of claim 5, wherein

determining the decimation group of frames comprises:

selecting a pitch decimation group of frames from the larger of the two arrays, the pitch decimation group being the selection of frames which, if decimated, yields the best alignment between the pitch values of the two arrays after decimation;

selecting an energy decimation group of frames from the larger of the two arrays, the energy decimation group being the selection of frames which, if decimated, yields the best alignment between the energy values of the two arrays after decimation;

selecting a plurality of formant coefficient decimation groups, each formant coefficient decimation group being a selection of frames from the larger of the two arrays which, if decimated, yields the best alignment between the formant coefficient values of the two after decimation; and

selecting frames from the larger of the two arrays for the decimation group by weighted average from the pitch decimation group, the energy decimation group, and each formant coefficient decimation group.

24. (New Claim) The speaker verification server of claim 10, wherein the verification application performs time domain normalization by:

comparing the quantity of frames in the array with the quantity of frames in the sample array to determine the quantity of frames to be decimated from the larger of the two arrays such that the two arrays are of an equal quantity of frames;

selecting a plurality of audio formant decimation groups, each audio formant decimation group being a selection of frames from the larger of the two arrays which, if decimated, yields the best alignment between a formant coefficient value of each frame of each the array and the corresponding formant coefficient value of each frame of the sample array after decimation; and

determining a decimation group of frames from the larger of the two arrays, the decimation group being a quantity of frames equal to the quantity of frames to be decimated and being the frames which are selected by weighted average from each of the audio formant decimation groups; and

decimating the decimation group of frames from the larger of the two arrays.